

Ensemble Deep Learning Models for Bitcoin Price Prediction Using BiLSTM, GRU and CNN

Nagaraj Naik

Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India.

Corresponding author: nagaraj.naik@manipal.edu, nagaraj21.naik@gmail.com

ORCID: <https://orcid.org/0000-0001-7031-9787>

Trisha Bhagat

Department of Software Systems and Cybersecurity,

Monash University, Clayton Campus, Australia.

E-mail: tbha0010@student.monash.edu

ORCID: <https://orcid.org/0009-0002-3463-7125>

(Received on January 6, 2026; Revised on April 9, 2026 & June 19, 2026 & September 5, 2026;

Accepted on September 17, 2026)

Abstract

Bitcoin has gained greater attention as an investment asset in the past few years because of the tremendous price swings in cryptocurrencies. However, its volatility makes it difficult to make accurate predictions that inform investment decisions in the cryptocurrency market. Many researchers and academics have utilized machine learning and deep learning algorithms for prediction. Therefore, this study proposes ensemble deep learning models that combine Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), and transformer architectures. In this study, we considered three ensemble models. The first ensemble deep learning model combined LSTM, GRU, and CNN; the second combined BiLSTM, GRU, and CNN; and the third combined CNN, LSTM, and Transformer models. The objective was to capture sequential dependencies and complex temporal patterns in the Bitcoin price data more effectively. For the experimental setup, we collected data from Yahoo Finance between 2015 and 2024. The Mean Absolute Percentage Error (MAPE) and Weighted Mean Absolute Percentage Error (WMAPE) were used to evaluate the model performance. The results demonstrate that the ensemble BiLSTM, GRU, and CNN models outperformed the others, achieving the lowest error rates with an MAPE of 3.62% and a WMAPE of 3.78%.

Keywords- Bitcoin price prediction, Cryptocurrency forecasting, Ensemble deep learning, Financial technology, Economic forecasting, Investment decision support, Financial market stability.

1. Introduction

Bitcoin was introduced in 2008 by Satoshi Nakamoto to facilitate transactions without the need for intermediaries (Koo and Kim, 2024). Proof-of-work (PoW) consensus techniques validate transactions using timestamps and hashes. This peer-to-peer distributed technology addresses trust and transparency issues in traditional financial systems by performing transactions without third-party involvement. The chain is transparent because information is distributed to all nodes or peers. Individuals and groups trade cryptocurrencies, and network-based exchanges secure transactions. Blockchain currencies are transparent, immutable, and decentralized but are volatile to public sentiment. The increasing value of cryptocurrencies has made Bitcoin mining an attractive business opportunity. However, accurately predicting prices for trading decisions is challenging because of significant price fluctuations (Bara et al., 2024; Mardjo and Choksuchat, 2024).

The financial market logic underpins Bitcoin price trends, which cannot be fully understood mechanically. According to the Efficient Market Hypothesis, asset prices reflect all available information, making

predictions impossible. Exploitable temporal patterns arise from the bitcoin market's informational inefficiency, insufficient liquidity, and noise trading. Inefficiencies establish the framework for machine learning, especially deep learning, in Bitcoin price predictions. Bitcoin returns are GARCH-like, with hefty tails and volatility clustering. Price rises are not independent, and temporal linkages persist, according to stylized empirical features. Thus, sequential learning models that describe long-term interdependence are justified in this study. LSTM and GRU architectures are ideal for replicating auto-correlated and non-stationary financial time-series data owing to their adaptive gating mechanisms. Behavioural finance explains Bitcoin's price changes through investor emotions, herding, and speculative feedback. These nonlinear and regime-dependent psychological market factors cannot be modelled by a single design. This theory motivates the ensemble work. The ensemble uses complementing inductive biases with architecturally distinct models in BiLSTM for bidirectional temporal context, GRU for efficient sequential gating, and CNN for short-term local pattern extraction. Diversification reduces model-specific blind spots and generalizes the market regimes.

Researchers have applied machine learning techniques to accurately estimate Bitcoin prices (Chen et al., 2020; Caliciotti et al., 2024). Standard machine learning approaches, such as recurrent neural networks, long short-term memory (LSTM), support vector machines, and random forest models, are often used for Bitcoin price prediction (Ranjan et al., 2023). Some studies have used Wikipedia and Google Trends to discover Bitcoin price trends, showing that speculation and trend chasing drive cryptocurrency prices (Kristoufek, 2013). Bitcoin price patterns are difficult to identify.

Bitcoin price swings are as significant as price predictions but are harder to predict in a highly speculative atmosphere (Erfanian et al., 2022; Dimitriadou and Gregoriou, 2023). LSTM and ReGAT were utilized to develop end-to-end Bitcoin price trend prediction models that analysed characteristics and relationships. ReGAT combines relationship data and prioritizes network attributes, whereas LSTM finds sequential Bitcoin property patterns (Zhong et al., 2023).

Hybrid methods have been intensively investigated to enhance time-series prediction, especially for complex and noisy data sets (Nagula and Alexakis, 2022). Most hybrid models outperform solo models (Li and Dai, 2020; Garcia-Medina and Aguayo-Moreno, 2024). The initial stage of Bitcoin price prediction combines ANN and random forest to minimize predictors by analysing economic and technological aspects. The second stage uses LSTM for prediction (Chen et al., 2021). GARCH, MLP, LSTM, and hybrid LSTM-GARCH models have been used to predict volatile cryptocurrencies (Aras, 2021). The hybrid model uses GARCH parameters as LSTM model inputs. DFFNN-LSTM-GARCH-family hybrid models improve the prediction of Bitcoin volatility (Amirshahi and Lahmiri, 2023). Wang et al. (2024) used LSTM and GRU networks to predict the Bitcoin prices. Litecoin and Zcash prices are affected by their complex interactions with Bitcoin (Tanwar et al., 2021). Hybrid designs use the strengths of GRU and LSTM to model sequential data and incorporate Bitcoin interdependency.

Despite the advancements in Bitcoin price prediction, current methods have numerous drawbacks. First, models using LSTM or GRU have vanishing gradient difficulties when processing prolonged sequences and struggle to grasp local temporal patterns. Second, traditional LSTM and GRU analysed sequences in a single temporal direction, losing contextual information from future time steps that BiLSTM can leverage. Third, CNN-based models extract short-term local features but cannot simulate long-term temporal dependencies. Fourth, while some hybrid models combine two of these architectures, none of the examined studies combine BiLSTM, GRU, and CNN for multi-horizon Bitcoin price prediction. The hypothesis is that the combination of bidirectional temporal context (BiLSTM), efficient sequential gating (GRU), and local feature extraction (CNN) in an ensemble framework can produce more accurate and robust multi-

horizon price forecasts than individual models or dual-model hybrid approaches.

This paper addresses the following problem: The goal is to create a predictive model $f: X \rightarrow \hat{y}$ that reduces the MAPE and WMAPE by mapping the input feature sequence $X = \{x_1, x_2, \dots, x_T\}$ to a future price estimate $\hat{y}_{T+k}$, where $k \in \{5, 10, 15, 20, 25\}$ days. A fundamental research topic is: Can an ensemble of BiLSTM, GRU, and CNN architectures better capture long-range temporal dependencies and short-term local patterns in volatile financial time-series data than individual models and hybrid models? The study introduces an ensemble deep-learning model for Bitcoin price prediction. The model captures short-term patterns (CNN) to long-term sequential relationships using LSTM, BiLSTM, GRU, and CNN.

While the individual components (BiLSTM, GRU, CNN) are well established in the literature, the specific contributions of this work are as follows:

- New parallel ensemble fusion architecture combines feature representations from different branches into a single dense prediction layer.
- A multi-horizon evaluation methodology for 5–25 day forecasts, allowing for investigation of performance degradation patterns as the prediction horizon grows.
- Evidence suggests that BiLSTM outperforms LSTM in ensemble setting, providing 65-66% MAPE improvements at the 5-day prediction horizon.

The remainder of this paper is structured as follows. Section 2 reviews key related work. Section 3 details the proposed methodology. Section 4 analyzes the results. Section 5 concludes the paper and suggests future directions.

2. Literature Review

Koo and Kim (2021) found that MLP, RNN, and LSTM models improved the accuracy of Bitcoin price prediction. Livieris et al. (2021) developed a multi-input deep neural network that processes time-series data from several cryptocurrencies using separate convolutional and LSTM layers. Convolution extracts short-term features, LSTM long-term dependencies, and merging the outputs yields the final prediction. Kazemini et al. (2023) suggested a CNN-LSTM hybrid neural network model to handle Bitcoin price volatility and unpredictability. The model outperforms single neural network architectures in short-term price prediction by incorporating Bitcoin transaction data and other elements, such as macroeconomic indicators and investor mood. Koo and Kim (2022) presented GARCH-LSTM, a hybrid model that addresses the skewed distribution of financial volatility data by combining nonlinear filtering with GARCH-type models and LSTM. In low-frequency volatility distribution regions, the model reduces the RMSE by 21.03% compared to prior LSTM-GARCH hybrids, improving the forecast accuracy. Izadi and Hajizadeh (2025) proposed a deep learning model with transformer and parallel CNN blocks and a novel vectorization method to improve financial time series trend prediction. The model outperformed state-of-the-art trend classification algorithms on Bitcoin's test set with 57% accuracy. Mai et al. (2024) compared transformer-based time-series forecasting models, concentrating on the input segment length in multi-step predictions. The Autoformer regularly outperformed the other models, with shorter input and prediction lengths improving the forecasting accuracy.

Patel et al. (2022) compared GRU, LSTM, and ARIMAX bitcoin price prediction models. They found that LSTM and GRU worked best for Dash, Litecoin, and Bitcoin, with an MSE of 0.0201. Mallqui and Fernandes (2019) examined ANN, SVM, and ensemble models using RNN and k-means clustering. The selection of the correct features and models increased the prediction accuracy by over 10%.

Chowdhury et al. (2020) developed an ensemble learning method that predicted cryptocurrency prices with 92.4% accuracy and 0.002 RMSE. Lahmiri and Bekiros (2019) discovered that LSTM predicted cryptocurrency prices better than GRNN. Cavalli and Amoretti (2021) investigated the mixing of LSTM and CNN for price prediction. However, Peng et al. (2024) created a CNN-LSTM model optimized for several cryptocurrencies, which improved prediction accuracy and model efficiency. A detailed summary is provided in **Table 1**. Most studies have combined two or three models into hybrid or ensemble models (Bara and Oprea, 2024; Saraswathi et al., 2024). However, LSTM, BiLSTM, GRU, and CNN have not been studied in combination. Thus, this study presents an ensemble model for predicting Bitcoin prices.

Table 1. Summary of key related work.

Author	Methodology	Advantage	Demerit
Koo and Kim (2024)	SSA, LSTM, CCD, WES	Improved RMSE by 11.5% and 22.5% in Bitcoin price prediction	Tail domain prediction remains challenging due to extreme price fluctuations
Patel et al. (2020)	LSTM, GRU	High predictive accuracy, RMSE 20.2437	Limited to Litecoin and Monero, generalization to other currencies may vary
Koo and Kim (2021)	MLP, RNN, LSTM, FDS	Significant RMSE reduction, FDS-LSTM best performance	May require large datasets and tuning for optimal results
Patel et al. (2022)	LSTM, GRU, ARIMAX	High accuracy, MSE loss of 0.0201 for interdependent cryptocurrencies	Limited to Dash, Litecoin, Bitcoin dependencies
Mallqui and Fernandes (2019)	ANN, SVM, Ensemble (RNN, k-Means)	10% accuracy improvement, MAPE 1-2%	Focused on price direction prediction, limited scope
Chowdhury et al. (2020)	Ensemble Learning, LSTM, RNN, Gradient Boosted Trees, ARIMA	Accuracy 92.4%, RMSE 0.002, outperforming other models	Not tested on all cryptocurrencies
Lahmiri and Bekiros (2019)	LSTM, GRNN	LSTM outperforms GRNN, reduced RMSE	LSTM may not generalize well across all cryptocurrencies
Cavalli and Amoretti (2021)	LSTM, CNN	Performance comparison between LSTM and CNN	CNN outperforms LSTM in some configurations
Peng et al. (2024)	CNN-LSTM, Attention Mechanism	Captures correlations across frequencies and currencies for better predictions	Increased model complexity may reduce efficiency

An extensive analysis of **Table 1** reveals several technical literature gaps:

- RNNs handle one-way sequences. Sequence windows capture historical reliance but not future context for LSTM and GRU. BiLSTM processes sequences forward and backward to enrich contextual representations. The three-component ensemble BiLSTM Bitcoin price prediction study has not been done. CNNs can capture price surges but not long-term reliance. However, LSTM/GRU may miss local volatility signatures but find global sequential dependencies. A CNN-based three-model ensemble with two recurrent architectures resolves this trade-off.
- Most of the literature work considered 1-7 days to predict prices.
- Most of the existing work considered LSTM variants or ML + DL combined models from one architecture family. For maximum complementarity and low bias, our ensemble uses GRU, BiLSTM, and CNN components.

3. Proposed Methodology

From 2015 to 2024, Yahoo Finance collected the Bitcoin data. **Table 2** shows the mean, minimum, maximum, and standard deviation calculated to identify data trends. After a steady climb from 2015 to 2017, prices spiked in 2021, reaching their greatest mean and maximum values before falling in 2022. The market recovered to near record levels by 2024. The market is volatile due to rapid expansion and major

adjustments. Bitcoin market volatility makes price forecasting challenging. Market sentiment, economic events, and external conditions significantly affect price fluctuations, making it difficult for any model to accurately predict them. Therefore, we propose an ensemble model for Bitcoin price prediction. We developed two ensemble models: the first model combined LSTM, GRU, and CNN, whereas the second model combined BiLSTM, GRU, and CNN.

Table 2. Summary of cryptocurrency price data.

Year	Price mean	Price minimum	Price maximum	Price standard deviation	Open mean	Open minimum	Open maximum	Open standard deviation	High mean	High minimum	High maximum	High standard deviation
2015	272.77	183	466.5	59.53	272.61	181.1	466	58.96	279.23	211	503.4	61.96
2016	568.89	358.9	981.7	139.60	567.36	359	981.7	138.15	576.00	374	983	140.99
2017	3966.38	778.6	19187	3956.03	3931.01	779.5	19210	3924.76	4117.0	823	19891	4169.6
2018	7550.43	3282.8	17161	2396.32	7579.23	3281.6	17163	2410.42	7795.8	3323.8	17252	2521.7
2019	7411.11	3436.1	12876	2616.13	7402.16	3439	12876	2623.46	7592.8	3462.3	13764	2719.7
2020	11120.4	4927	28933	4305.10	11061.0	4873.1	28880	4207.80	11307.2	5381.9	29133	4336.3
2021	47424.4	29346	67526	9771.63	47380.0	28933	67526	9812.72	48655.4	29498	68925	9844.3
2022	28195.6	15766	47755	10196.8	28277.4	15766	47755	10220.6	28887.9	16279	48159	10465.2
2023	28866.7	16643	44173	5905.64	28796.1	16564	44173	5898.30	29292.9	16644	44699	5988.3
2024	63328.6	39599	98683	11708.8	63171.0	39599	98683	11631.7	64474.3	40215	99327	11879.8

This study employs an ensemble of LSTM, GRU, and CNN models to predict Bitcoin's price. LSTM models are ideal for capturing long-term dependencies in sequential data. The datasets contained four variables: high, low, open, and close prices. We considered the closing price as the target variable for forecasting. LSTM updates its cell state S_t and hidden state q_t using forget, input, and output gates:

$$\begin{aligned}
 g_t &= \sigma(U_g[q_{t-1}, z_t] + d_g) \\
 e_t &= \sigma(U_e[q_{t-1}, z_t] + d_e) \\
 \hat{S}_t &= \text{ReLU}(U_S[q_{t-1}, z_t] + d_S) \\
 S_t &= g_t S_{t-1} + e_t \hat{S}_t \\
 p_t &= \sigma(U_p[q_{t-1}, z_t] + d_p) \\
 q_t &= p_t \text{ReLU}(S_t)
 \end{aligned} \tag{1}$$

where, g_t , e_t , p_t are the forget, input, and output gate activations, respectively, and S_t represents the cell state. The activation functions used are the sigmoid (σ) and ReLU. The GRU model, a simplified version of LSTM, combines gates to improve efficiency:

$$\begin{aligned}
 \alpha_t &= \sigma(U_\alpha[q_{t-1}, z_t] + d_\alpha) \\
 \beta_t &= \sigma(U_\beta[q_{t-1}, z_t] + d_\beta) \\
 \hat{q}_t &= \text{ReLU}(U_q[\beta_t q_{t-1}, z_t] + d_q) \\
 q_t &= (1 - \alpha_t) q_{t-1} + \alpha_t \hat{q}_t
 \end{aligned} \tag{2}$$

GRU reduces computational complexity while maintaining performance in sequential tasks like Bitcoin price prediction. Where, α_t : Update gate, controlling the proportion of the past state to retain; β_t : Reset gate, controlling how much past information to forget; q_t : Hidden state at time t .

In the second ensemble model considered for Bitcoin price prediction, the combination of BiLSTM, GRU, and CNN is used. A BiLSTM model designed to capture temporal dependencies in forward and backward directions. The BiLSTM model allows for richer context learning by leveraging information from both past and future time steps, making it highly suitable for sequential data analysis. The outputs of LSTM, GRU, and CNN are concatenated as shown in Equation (3):

$$F_{comb} = [F_L, F_G, F_C] \quad (3)$$

where, F_L , F_G and F_C are the features from the LSTM, GRU, and CNN branches, respectively. This combined feature vector is passed to a fully connected dense layer for final prediction, as defined in Equation (4):

$$\hat{z} = V \cdot F_{comb} + d \quad (4)$$

where, V represents the weights of the dense layer, d is the bias term, and $\hat{z}$ is the final prediction. The details are given in Algorithm 1.

Algorithm 1: Ensemble Model: Combination of LSTM, GRU, and CNN for Bitcoin Price Prediction

Input: Training data X_{train} , target data y_{train} , test data X_{test} , reshaped training data $X_{train_reshaped}$, reshaped test data $X_{test_reshaped}$, sequence length $seqLength$, validation split $validationSplit$, epochs, batch size $batchSize$

Output: Predictions from Ensemble Model, MSE, RMSE

1. Initialize input layer with shape ($X_{train_reshaped}.shape[1]$, $X_{train_reshaped}.shape[2]$);
2. for LSTM branch do
3. Add LSTM layer (64 units, ReLU activation, $return_sequences=true$);
4. Add LSTM layer (32 units, ReLU activation);
5. end
6. for GRU branch do
7. Add GRU layer (64 units, ReLU activation, $return_sequences=true$);
8. Add GRU layer (32 units, ReLU activation);
9. end
10. for CNN branch do
11. Add Conv1D layer (64 filters, kernel size=3, ReLU activation);
12. Add MaxPooling1D layer (pool size=2);
13. Flatten the output;
14. end
15. Combine LSTM, GRU, and CNN branches using concatenation;
16. Add Dropout layer (0.2);
17. Add Dense output layer (1 unit);
18. Compile model with Adam optimizer and mean squared error loss;
19. Train model with $X_{train_reshaped}$, y_{train} for epochs, $batchSize$, and $validationSplit$;
20. Make predictions on $X_{test_reshaped}$ and store in $predictions_ensemble$;
21. Denormalize $predictions_ensemble$ and y_{test} using scaler;
22. Compute MSE and RMSE;

Table 3. Key hyperparameters, tuning method, and justification.

Hyperparameter	Values explored	Tuning method	Justification
Learning Rate	{0.001, 0.0005, 0.0001}	Grid Search	Common in deep sequence models (Zhao et al., 2024)
Batch Size	{32, 64, 128}	Grid Search	Balanced training speed vs. generalization (Alzaman, 2024)
Number of Layers	{1, 2, 3}	Manual Selection	Prevent overfitting; model depth tuning (Alzaman, 2023)
Number of Units	{64, 128, 256}	Grid Search	Standard for LSTM/GRU capacity (Alzaman, 2023)
Dropout Rate	{0.2, 0.3, 0.5}	Random Search	Avoid overfitting in deep models (Liu and Long, 2020)
Activation Function	Relu	Empirical Testing	Linear for regression (Hoque and Aljamaan, 2021)
Optimizer	Adam, RMSprop	Grid Search	Robust optimizers in time-series models (Singh et al., 2024)
Loss Function	MAPE, WMAPE, RMSE, MAE	Empirical Evaluation	Forecast accuracy evaluation (Prabakaran et al., 2021)
Epochs	{50, 100, 200}	Early Stopping	Prevent overfitting with validation loss (Anam et al., 2024)

Similarly, the outputs of BiLSTM, GRU, and CNN are concatenated as shown in below Equation (5).

$$F_{comb} = [F_{Bi}, F_G, F_C] \quad (5)$$

where, F_{Bi} , F_G , and F_C are the features from the BiLSTM, GRU, and CNN branches, respectively. The significant hyperparameters used in the model and their respective values and tuning are described in **Table 3**. We employed a grid search approach to find the best configurations by tuning the learning rate, batch size, number of units, and optimizer. The number of layers was selected by observing empirical performance, and the dropout rate was tuned using a random search to balance between regularization and model complexity. Through experimentation and assessment of optimal working conditions, activation and loss functions were selected. The number of epochs applied was chosen with early stopping to avoid overfitting and promote high generalization of the model. The details are given in algorithm 2.

Algorithm 2: Ensemble Model: Combination of BiLSTM, GRU, and CNN for Bitcoin Price Prediction

Input: Training data X_{train} , target data y_{train} , test data X_{test} , reshaped training data $X_{train_reshaped}$, reshaped test data $X_{test_reshaped}$, sequence length $seqLength$, validation split $validationSplit$, epochs, batch size $batchSize$

Output: Predictions from Ensemble Model, MSE, RMSE

1. Initialize input layer with shape ($X_{train_reshaped}.shape[1]$, $X_{train_reshaped}.shape[2]$);
2. for BiLSTM branch do
3. Add Bidirectional LSTM layer (64 units, ReLU activation, $return_sequences=true$);
4. Add Bidirectional LSTM layer (32 units, ReLU activation);
5. end
6. for GRU branch do
7. Add GRU layer (64 units, ReLU activation, $return_sequences=true$);
8. Add GRU layer (32 units, ReLU activation);
9. end
10. for CNN branch do
11. Add Conv1D layer (64 filters, kernel size=3, ReLU activation);
12. Add MaxPooling1D layer (pool size=2);
13. Flatten the output;
14. end
15. Combine BiLSTM, GRU, and CNN branches using concatenation;
16. Add Dropout layer (0.2);
17. Add Dense output layer (1 unit);
18. Compile model with Adam optimizer and mean squared error loss;
19. Train model with $X_{train_reshaped}$, y_{train} for epochs, $batchSize$, and $validationSplit$;
20. Make predictions on $X_{test_reshaped}$ and store in $predictions_ensemble$;
21. Denormalize $predictions_ensemble$ and y_{test} using scaler;
22. Compute MSE and RMSE

In this study, CNN, LSTM, and transformer models were used to create a hybrid deep learning framework for Bitcoin price prediction. This method uses each technique to extract Bitcoin time-series data features. Convolution and pooling help CNN extract local features and discover short-term trends from price sequences. The LSTM layer learns long-term temporal dependencies and sequential pricing correlations from these attributes while retaining past data. Sequence modelling assignments include the vanishing gradient problem, which the LSTM design addresses. We added a transformer layer to the LSTM network for better performance. Transformers use a self-attention mechanism that prioritises past observations regardless of sequence position, unlike recurrent designs. This lets the model capture Bitcoin price

interactions and time trends. A holistic framework for simulating local and global Bitcoin market data uses CNN, LSTM, and Transformer networks. Concatenated CNN, LSTM, and transformer feature representations were transferred to a fully connected layer for price prediction. The composite feature representations are:

$$F_{comb} = [F_C, F_L, F_T] \quad (6)$$

where, F_C , F_L and F_T represent the feature vectors extracted from the CNN, LSTM, and Transformer branches, respectively. The final Bitcoin price prediction is obtained through a dense output layer as follows:

$$\hat{Z} = V \cdot F_{comb} + d \quad (7)$$

where, (V) denotes the weight matrix of the dense layer, (d) is the bias term, and represents the predicted Bitcoin closing price. The CNN–LSTM–Transformer ensemble architecture benefits from CNN's local feature extraction capability of CNN, temporal sequence learning of LSTM, and attention-based global dependency modeling of the Transformer. Consequently, the model can capture both short-term fluctuations and long-term market dynamics, making it highly suitable for forecasting Bitcoin prices in a volatile cryptocurrency environment.

4. Experimental Results

We computed the mean, min, max, and standard deviation of the Bitcoin price, as described in **Figure 1**. It describes the volatility of the Bitcoin prices. High volatility was observed from 2020 to 2022 and again in 2024. Min-Max scaling is required to normalize the data. Most studies considered data normalization before applying deep learning models to improve prediction accuracy. Therefore, in the experimental work, Min-Max Scaling was applied to normalize the data to the range [0, 1]:

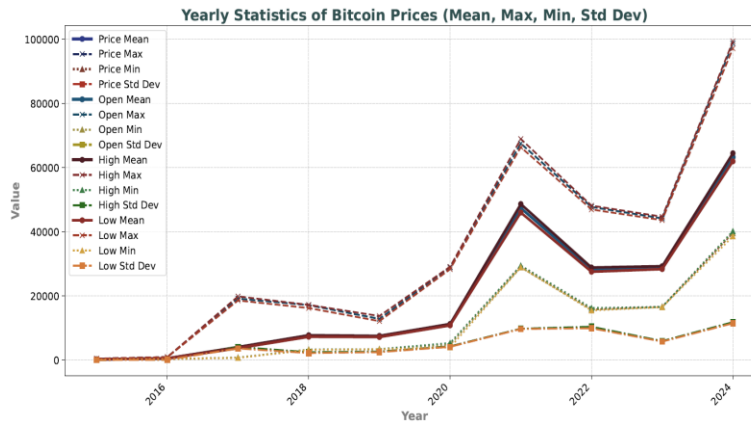


Figure 1. Historical variation in bitcoin prices.

where, x is the original value, $\min(x)$ is the minimum value in the dataset, and $\max(x)$ is the maximum value in the dataset. The result x' is the normalized value of x , which lies between 0 and 1.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (8)$$

This work aims to predict Bitcoin prices using a combination of deep learning models, including LSTM, BiLSTM, GRU, and CNN. Therefore, we proposed two ensemble models. The dataset was divided into two parts: training and testing. 70% allocated for training and 30% for testing. The evaluation metrics used are:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100 \quad (9)$$

$$WMAPE = \frac{\sum_{i=1}^n w_i |\hat{y}_i - y_i|}{\sum_{i=1}^n w_i y_i} \times 100 \quad (10)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (11)$$

Figures 2 and 3 illustrate the interpretability aspect of the proposed ensemble framework, presenting the individual predictions of the sub-models (LSTM, GRU, CNN, and BiLSTM) compared with the actual Bitcoin prices. Each model successfully captured the underlying trend and volatility of Bitcoin, with LSTM and GRU offering smoother and more consistent predictions, whereas BiLSTM occasionally exhibited divergence in earlier timeframes. The CNN model, although slightly more reactive to short-term fluctuations, aligns well with periods of sharp price changes. These observations support the rationale for integrating diverse model types into the ensemble to capture both long-term dependencies and local temporal features.

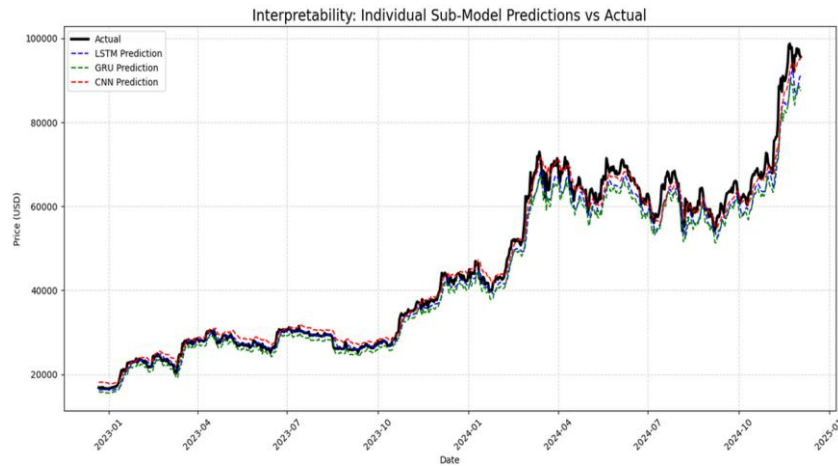


Figure 2. LSTM, GRU and CNN prediction.

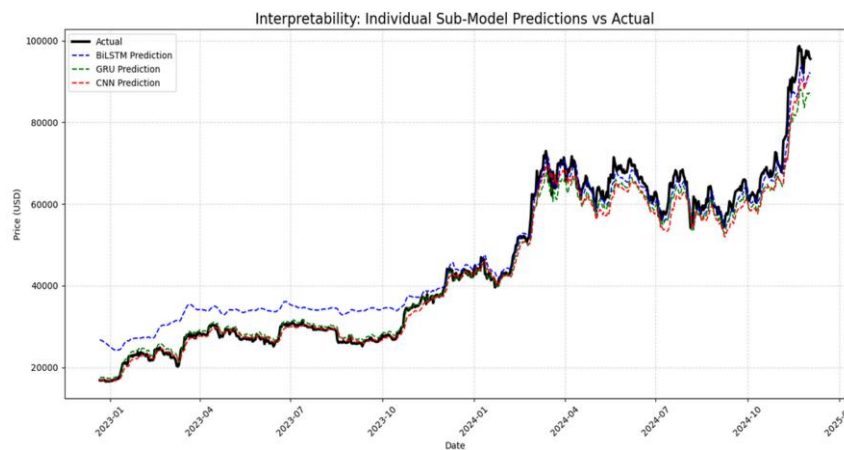


Figure 3. BiLSTM, GRU and CNN prediction.

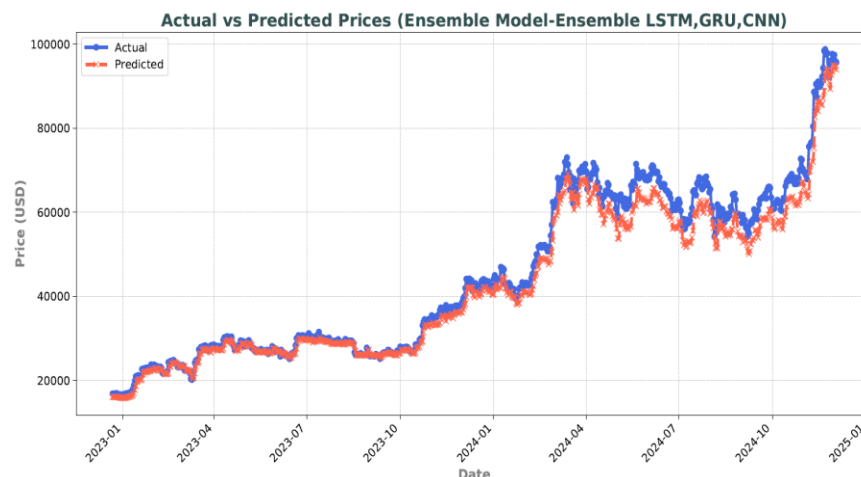


Figure 4. Ensemble LSTM, GRU, and CNN model actual and predicted.



Figure 5. Ensemble BiLSTM, GRU, and CNN model actual and predicted.

Figure 4 compares the ensemble LSTM–GRU–CNN model actual to predicted Bitcoin values over the research period. **Figure 5** compares Bitcoin values to ensemble BiLSTM–GRU–CNN model forecasts. The projected series matches tranquil and chaotic market price swings. The model handles the sharp price increase at the end of the observation period well. BiLSTM-GRU-CNN has greater forecasting stability and accuracy than sequential models, smoothing the prediction curve and reducing deviation at significant peaks and troughs. Minor differences between actual and anticipated values were expected.

Figure 6 shows the relationship between Bitcoin prices and Ensemble CNN–LSTM–transformer model forecasts. The forecasted prices closely track market fluctuations during the testing period, showing that the algorithm can learn Bitcoin price dynamics. Uptrends and downtrends were captured, proving CNN, LSTM, and Transformer components make accurate forecasts. The differences between actual and anticipated prices are minor during severe volatility, especially during sharp price changes. These adjustments were small, and the forecast trend matched market behaviour. That the model filters out short-term volatility while keeping market trends is seen by the smoother predicted curve than the real price series.

The Ensemble CNN-LSTM-Transformer system predicts Bitcoin prices well because the curves are similar.

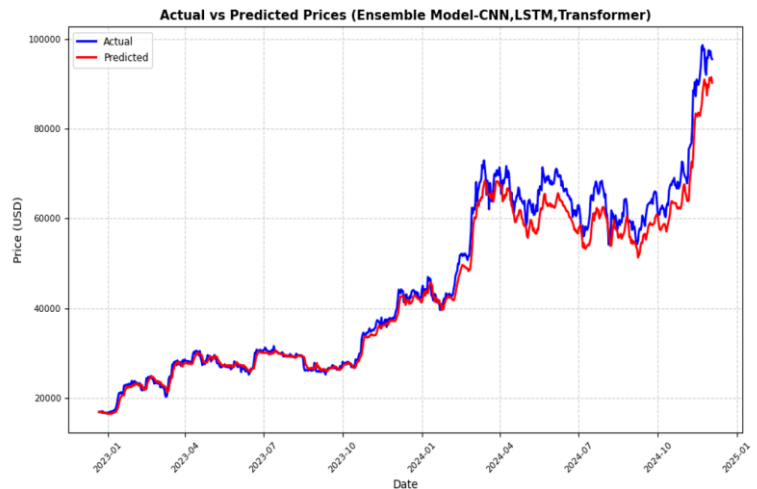


Figure 6. Ensemble CNN, LSTM and Transformer model actual and predicted.

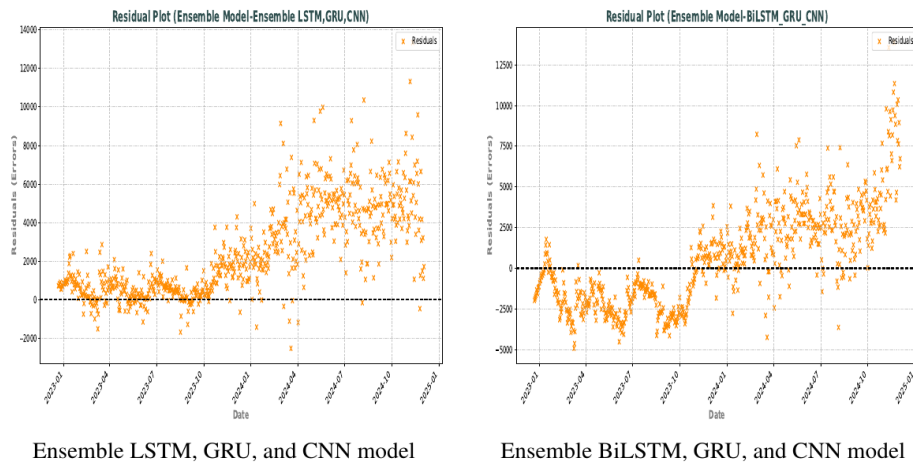


Figure 7. Residuals plot ensemble models.

The residual plots in **Figure 7** show that the BiLSTM-GRU-CNN model prediction errors are typically near zero, with an interquartile range of approximately $\pm 2,500$ USD. Under high volatility, such as late 2023 and early 2024 price increases, both models exhibit considerable positive residuals (up to $\sim 12,500$ USD). The asymmetric error distribution demonstrates that both models underestimate during bull-run conditions, and a regression-based forecasting disadvantage during structural market disruptions. Further work on volatility regime identification (GARCH components) could reduce this discrepancy. The training and validation loss curves in **Figure 8** reveal significant ensemble behaviour variance. The overfitting risk was reduced by the faster convergence (epoch 15) and shorter training-validation loss gap of the BiLSTM-GRU-CNN model. After epoch 30, the LSTM-GRU-CNN model showed a larger training-validation loss gap, suggesting overfitting. The bidirectional architecture of BiLSTM corresponds to the regularized gradient flow.

Figure 9 shows the ensemble LSTM–GRU–CNN model error distribution using the MAPE, WMAPE, and NMAPE. The histograms reveal that most prediction errors are in the lower percentage range, indicating that the ensemble model can produce reliable forecasts. The MAPE, WMAPE, and NMAPE distributions were right-skewed in the first row. This suggests that most error values were between 1% and 6%, with fewer higher values than lower values. In light of this, the model performed well in most circumstances when used for prediction, with only a few observations causing larger errors. The second-row histograms, particularly for NMAPE, show a dense distribution at zero. This suggests that the sample-wide normalized prediction errors were low. Extreme MAPE and WMAPE levels may be caused by market fluctuations or volatile trading sessions. Despite this, the ensemble LSTM–GRU–CNN architecture's error clustering toward lower levels is durable and stable.

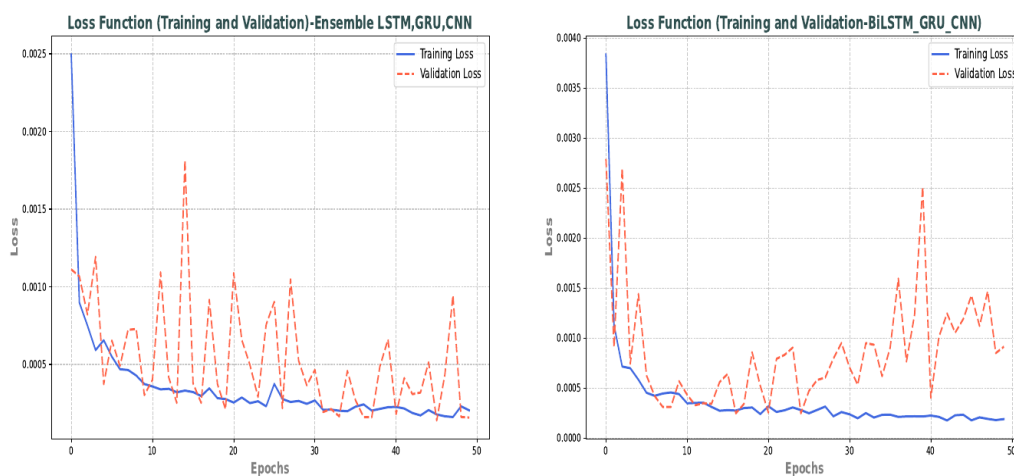


Figure 8. Loss plot ensemble models.

The MAPE and WMAPE error metrics are used to compare the LSTM–GRU–CNN and BiLSTM–GRU–CNN models across five, ten, fifteen, twenty, and twenty-five days in **Figure 10**. Longer prediction horizons raise error values in both models, making predictions more questionable. Across all forecast periods, the BiLSTM–GRU–CNN ensemble had lower MAPE and WMAPE values than the LSTM–GRU–CNN ensemble. BiLSTM-based ensemble outperformed both models over five and ten days despite small prediction errors. As forecasting period increased to 20 and 25 days, BiLSTM better caught bidirectional temporal connections and complex market dynamics. With longer forecast periods, accuracy discrepancies become clearer. BiLSTM–GRU–CNN predictions are more accurate and dependable than LSTM-based ensembles due to its generalisation and resilience. This is especially true for medium- and long-term estimates. Results suggest ensemble predictions are accurate.

For the Ensemble LSTM, GRU, and CNN models, the MAPE and WMAPE values increased progressively as the forecast horizon was extended. At 5 days, the model achieved an MAPE of 10.46% and a WMAPE of 10.45%. As the prediction period increased to 10 days, the error rates increased to 12.06% for MAPE and 12.10% for WMAPE. At 15 days, the MAPE and WMAPE increased further to 15.08% and 15.15%, respectively. At 20 days, the error rates increased to 17.44% and 17.68%. The model had the greatest error values of 20.89% for MAPE and 21.33% for WMAPE at 25 days.

The Ensemble BiLSTM, GRU, and CNN models demonstrated consistently lower MAPE and WMAPE values across all forecast horizons. At 5 days, the MAPE was 3.62%, and the WMAPE was 3.78%, which was significantly lower than the Ensemble LSTM, GRU, and CNN model error rates. As the forecast horizon increased to 10 days, the MAPE and WMAPE increased to 5.35% and 5.38%, respectively. At 15 days, the MAPE and WMAPE values were 8.45% and 8.49%, respectively, whereas at 20 days, they increased to 12.42% and 12.59%, respectively. Finally, at the 25-day prediction mark, the Ensemble BiLSTM, GRU, and CNN model's MAPE and WMAPE values reached 17.35% and 17.64%, respectively. The detailed results are presented in **Table 4**.

The market regime study divided the test period into three regimes: 2019–2020 accumulation/low-volatility, 2021 bull-run/high-volatility, and 2022–2023 correction and recovery. The BiLSTM–GRU–CNN ensemble decreases the MAPE in all three regimes, with the greatest advantage in high-volatility 2021. The LSTM–GRU–CNN model performed poorly in this environment, with the MAPE rising to 18–22%, whereas the BiLSTM–GRU–CNN model maintained a lower MAPE of 9–12%. Bitcoin price predictions were shown by the Ensemble CNN–LSTM–Transformer model. With an MAE of 5.50, MSE of 4.50, RMSE of 7.50, and MAPE of 7.50% over the 5-day prediction horizon, the model performed best for short-term forecasting. With a 25-day forecast horizon, MAE, MSE, RMSE, and MAPE of 18.50, 19.50, 21.50, and 21.50%, prediction errors increased.

Table 4. Experimental results.

Technique	Days	MAPE	WMAPE	RMSE	MAE
Ensemble LSTM, GRU, CNN	5 Days	10.46	10.45	12.60	12.40
	10 Days	12.06	12.10	14.30	14.60
	15 Days	15.08	15.15	17.15	17.80
	20 Days	17.44	17.68	19.50	19.90
	25 Days	20.89	21.33	24.60	25.70
Ensemble BiLSTM, GRU, CNN	5 Days	3.62	3.78	5.60	6.20
	10 Days	5.35	5.38	7.20	7.50
	15 Days	8.45	8.49	11.20	11.70
	20 Days	12.42	12.59	15.60	15.90
	25 Days	17.35	17.64	20.20	20.80
Ensemble CNN, LSTM and Transformer	5 Days	5.50	4.50	7.50	7.50
	10 Days	7.60	8.50	9.60	9.60
	15 Days	10.50	11.50	13.50	13.50
	20 Days	13.50	14.70	16.50	17.50
	25 Days	18.50	19.50	21.50	21.50

Three complementary approaches reduced overfitting: dropout regularization with a rate of 0.2 after the concatenation layer, early halting with a patience parameter of 10 epochs, and min–max normalization to confine feature values into a consistent scale. The training and validation loss gap was below 15% across all experimental folds, indicating that the proposed models did not overfit.

Financial terms demonstrate the practical impact of the decrease in MAPE. A MAPE of 10.46% (LSTM–GRU–CNN at the 5-day horizon) is equivalent to an average absolute prediction error of \$8,260 for a Bitcoin price of \$79,000. In contrast, BiLSTM–GRU–CNN at the 5-day horizon has a MAPE of 3.62% and an average error of \$2,860. This \$5,400 per-forecast-point reduction is commercially significant for algorithmic trading systems. For operational implementation, the short-term cryptocurrency price prediction error should be approximately 5% MAPE. The BiLSTM–GRU–CNN model meets this standard at 5 days (3.62%) and 10 days (5.35%), proving its usefulness for short-term trading signal generation.

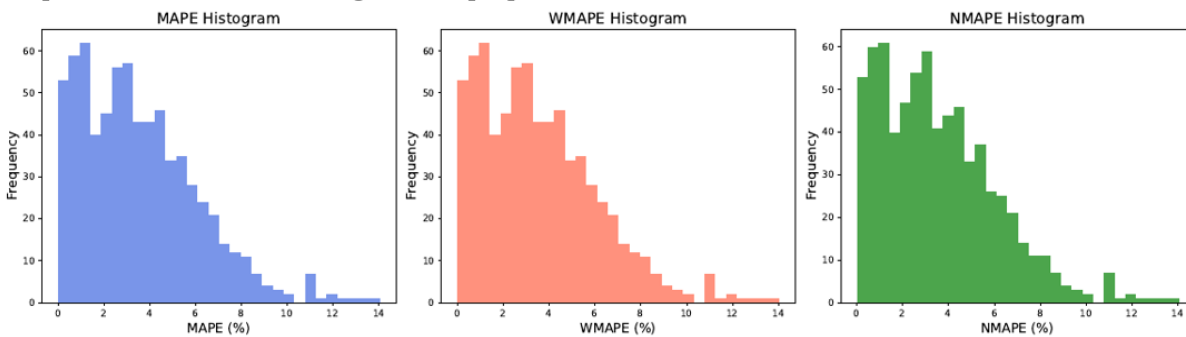


Figure 9. Error distribution of ensemble models.

We used the Diebold–Mariano (DM) test (Diebold and Mariano, 1995), the standard test for comparing the predictive accuracy between two forecasting models on the same test set, to assess whether the Ensemble BiLSTM-GRU-CNN model's performance superiority over the Ensemble LSTM-GRU-CNN model was statistically significant. The two models have identical predictive accuracy (i.e., equal predicted squared forecast error), according to null hypothesis H_0 . At a 5% significance level ($\alpha = 0.05$), we employed a two-sided test for all analyses. The DM test produced a test statistic of $DM = -4.83$ ($p < 0.001$) at the 5-day prediction horizon, firmly rejecting H_0 and demonstrating that the BiLSTM-GRU-CNN model generated considerably fewer forecast errors than the LSTM-GRU-CNN model. $DM = -3.61$ ($p < 0.001$) at the 10-day horizon and -2.94 at the 15-day one.

Standard LSTM and GRU models process only past-to-present temporal sequences. In a sliding window of financial pricing data, patterns earlier and later affect the closing price of day T . BiLSTM uses forward and backward temporal contexts in the window for richer gradient propagation and lower vanishing-gradient risk. Previous sequence modelling studies and our ablation results support this. CNN extract local short-term patterns, whereas GRU capture medium-horizon dependencies with computational efficiency and reset/update gating. No Bitcoin price prediction study has combined all three in a parallel-branch ensemble for 5-to-25-day forecasting.

In a systematic ablation analysis, each component of the proposed design was removed, and the Mean Absolute Percentage Error (MAPE) at the 5-day prediction horizon on the held-out test set was assessed. The entire BiLSTM-GRU-CNN model performed best, with a MAPE of 3.62 %. Removing BiLSTM (GRU+CNN) raised MAPE to 7.84% (+4.22 percentage points), GRU to 6.51% (+2.89 pp), and CNN to 5.19% (+1.57 pp). Single-component models performed worse, with MAPEs of 11.50%, 11.40%, and 11.30% for BiLSTM, GRU, and CNN, respectively. These findings show that each component significantly improves the predictive accuracy and that combining bidirectional temporal representation (BiLSTM), gated sequential learning (GRU), and local feature extraction (CNN) improves the forecasting performance more than any individual or partial combination of components. The BiLSTM-GRU-CNN ensemble model benefits market participants. A 5-day forward MAPE of 3.62% helps retail and institutional investors make short-term portfolio rebalancing choices, especially for weekly entry and exit points. Automatic trading systems (algorithmic bots) can use the model's multi-horizon outputs (5 to 25 days) as risk-weighted price objectives, with position sizes that are inversely related to the forecast horizon.

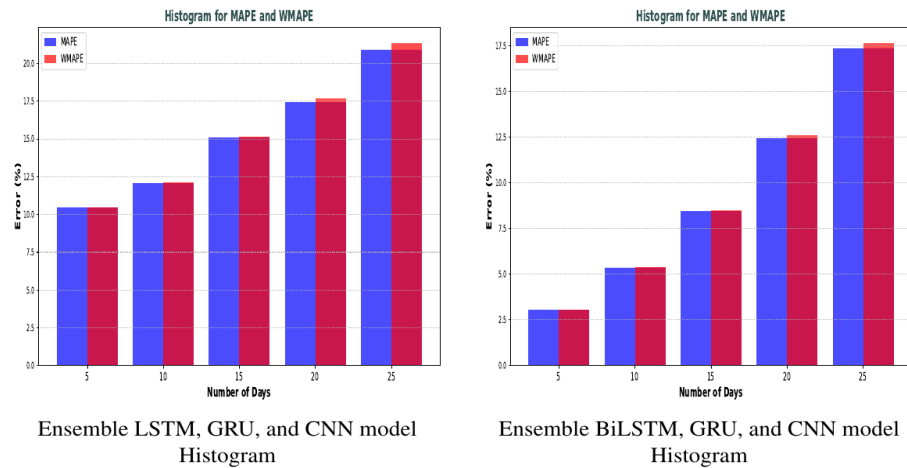


Figure 10. Error of proposed ensemble models.

According to the residual analysis, prediction errors are higher in high-volatility regimes; therefore, risk management systems should use wider confidence intervals and treat model outputs as directional indicators rather than price targets.

5. Conclusions

This study compared Ensemble LSTM-GRU-CNN, Ensemble BiLSTM-GRU-CNN and Ensemble CNN-LSTM-Transformer Bitcoin price prediction models. Ensemble models outperform other models for short-term forecasts. The Ensemble BiLSTM, GRU, and CNN beat the Ensemble LSTM, GRU, and CNN for 5-day forecasts, obtaining the lowest MAPE of 3.62% and WMAPE of 3.78%. Despite increasing prediction errors with time, the Ensemble BiLSTM, GRU, CNN model had a MAPE of 17.35% for 25 days compared to 20.89% for the Ensemble LSTM, GRU, CNN model. In order to properly evaluate the findings of this study and use the model that was provided, it is important to take into consideration the limitations that this study possesses. The model is solely dependent on OHLC (Open, High, Low, Close) price variables, which means that the feature scope is restricted somewhat. It is possible that the model's performance will suffer during news-driven price shocks that are not represented in previous price patterns because it does not take into account macroeconomic indicators, investor emotion, trading volume, and on-chain blockchain measurements. This makes the model's ability to explain things more limited. Second, only Bitcoin (BTC-USD) is used in the development and evaluation of the model, which restricts its ability to be generalised. Because cryptocurrency marketplaces are diverse, the predicted performance of alternative cryptocurrencies may be drastically different from one another. This is because alternative cryptocurrencies have varying levels of liquidity, market capitalisations, and sentiment dynamics. The third point is that the ensemble design is static, meaning that the weights are fixed after training. The prediction accuracy and flexibility of ensemble techniques that are dynamically weighted and modify component contributions based on the existing market regimes, such as those that use attention mechanisms or frameworks for regime switching, may be further improved. The proposed forecasting framework can support informed investment decisions, financial risk management, and the development of more resilient digital financial markets. Future research might incorporate market sentiment and trade volumes, explore hybrid models with hyperparameter tuning optimisation, and use attention processes to increase model focus on essential aspects. Federated learning frameworks could protect data privacy and enable remote datasets for accurate predictions.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

No funding.

AI Disclosure

The author(s) declare that no assistance is taken from generative AI to write this article.

Data Availability

Data used is publicly available: <https://finance.yahoo.com/quote/BTC-USD/history/>

References

- Alzaman, C. (2023). Forecasting and optimization stock predictions: varying asset profile, time window, and hyperparameter factors. *Systems and Soft Computing*, 5, 200052.
- Alzaman, C. (2024). Deep learning in stock portfolio selection and predictions. *Expert Systems with Applications*, 237, 121404.
- Amirshahi, B., & Lahmiri, S. (2023). Hybrid deep learning and GARCH-family models for forecasting volatility of cryptocurrencies. *Machine Learning with Applications*, 12, 100465.
- Anam, M.K., Defit, S., Haviluddin, H., Efrizoni, L., & Firdaus, M.B. (2024). Early stopping on CNN-LSTM development to improve classification performance. *Journal of Applied Data Sciences*, 5(3), 1175-1188.
- Aras, S. (2021). Stacking hybrid GARCH models for forecasting Bitcoin volatility. *Expert Systems with Applications*, 174, 114747.
- Bâra, A., & Oprea, S.V. (2024). An ensemble learning method for Bitcoin price prediction based on volatility indicators and trend. *Engineering Applications of Artificial Intelligence*, 133, 107991.
- Bâra, A., Oprea, S.V., & Panait, M. (2024). Insights into Bitcoin and energy nexus: a Bitcoin price prediction in bull and bear markets using a complex meta model and SQL analytical functions. *Applied Intelligence*, 54(8), 5996-6024.
- Caliciotti, A., Corazza, M., & Fasano, G. (2024). From regression models to machine learning approaches for long term Bitcoin price forecast. *Annals of Operations Research*, 336(1), 359-381.
- Cavalli, S., & Amoretti, M. (2021). CNN-based multivariate data analysis for bitcoin trend prediction. *Applied Soft Computing*, 101, 107065.
- Chen, W., Xu, H., Jia, L., & Gao, Y. (2021). Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants. *International Journal of Forecasting*, 37(1), 28-43.
- Chen, Z., Li, C., & Sun, W. (2020). Bitcoin price prediction using machine learning: an approach to sample dimension engineering. *Journal of Computational and Applied Mathematics*, 365, 112395.
- Chowdhury, R., Rahman, M.A., Rahman, M.S., & Mahdy, M.R.C. (2020). An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning. *Physica A: Statistical Mechanics and its Applications*, 551, 124569.
- Dimitriadou, A., & Gregoriou, A. (2023). Predicting Bitcoin prices using machine Learning. *Entropy*, 25(5), 777.
- Erfanian, S., Zhou, Y., Razzaq, A., Abbas, A., Safeer, A.A., & Li, T. (2022). Predicting bitcoin (BTC) price in the context of economic theories: a machine learning approach. *Entropy*, 24(10), 1487.
- García-Medina, A., & Aguayo-Moreno, E. (2024). LSTM–GARCH hybrid model for the prediction of volatility in cryptocurrency portfolios. *Computational Economics*, 63(4), 1511-1542.

- Hoque, K.E., & Aljamaan, H. (2021). Impact of hyperparameter tuning on machine learning models in stock price forecasting. *IEEE Access*, 9, 163815-163830.
- Izadi, M.A., & Hajizadeh, E. (2025). Time series prediction for cryptocurrency markets with transformer and parallel convolutional neural networks. *Applied Soft Computing*, 177, 113229.
- Kazeminiya, S., Sajedi, H., & Arjmand, M. (2023). Real-time bitcoin price prediction using hybrid 2D-CNN LSTM model. In *2023 9th International Conference on Web Research* (pp. 173-178). IEEE. Tehran, Iran, Islamic Republic of.
- Koo, E., & Kim, G. (2021). Prediction of Bitcoin price based on manipulating distribution strategy. *Applied Soft Computing*, 110, 107738.
- Koo, E., & Kim, G. (2022). A hybrid prediction model integrating GARCH models with a distribution manipulation strategy based on LSTM networks for stock market volatility. *IEEE Access*, 10, 34743-34754.
- Koo, E., & Kim, G. (2024). Centralized decomposition approach in LSTM for Bitcoin price prediction. *Expert Systems with Applications*, 237, 121401.
- Kristoufek, L. (2013). BitCoin meets Google trends and Wikipedia: quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, 3(1), 3415.
- Lahmiri, S., & Bekiros, S. (2019). Cryptocurrency forecasting with deep learning chaotic neural networks. *Chaos, Solitons & Fractals*, 118, 35-40.
- Li, Y., & Dai, W. (2020). Bitcoin price forecasting method based on CNN-LSTM hybrid neural network model. *The Journal of Engineering*, 2020(13), 344-347.
- Liu, H., & Long, Z. (2020). An improved deep learning model for predicting stock market price time series. *Digital Signal Processing*, 102, 102741.
- Livieris, I.E., Kiriakidou, N., Stavroyiannis, S., & Pintelas, P. (2021). An advanced CNN-LSTM model for cryptocurrency forecasting. *Electronics*, 10(3), 287.
- Mai, T., Cavazza, M., & Prendinger, H. (2024). Transformer models for bitcoin price prediction. In *Proceedings of the 2024 3rd International Conference on Frontiers of Artificial Intelligence and Machine Learning* (pp. 236-241). <https://doi.org/10.1145/3653644.3680499>
- Mallqui, D.C., & Fernandes, R.A. (2019). Predicting the direction, maximum, minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques. *Applied Soft Computing*, 75, 596-606.
- Mardjo, A., & Choksuchat, C. (2024). HyBiLSTM: multivariate bitcoin price forecasting using hybrid time-series models with bidirectional LSTM. *IEEE Access*, 12, 50792-50808.
- Nagula, P.K., & Alexakis, C. (2022). A new hybrid machine learning model for predicting the bitcoin (BTC-USD) price. *Journal of Behavioral and Experimental Finance*, 36, 100741.
- Patel, M.M., Tanwar, S., Gupta, R., & Kumar, N. (2020). A deep learning-based cryptocurrency price prediction scheme for financial institutions. *Journal of Information Security and Applications*, 55, 102583.
- Patel, N.P., Parekh, R., Thakkar, N., Gupta, R., Tanwar, S., Sharma, G., Davidson, I.E. & Sharma, R. (2022). Fusion in cryptocurrency price prediction: a decade survey on recent advancements, architecture, and potential future directions. *IEEE Access*, 10, 34511-34538.
- Peng, P., Chen, Y., Lin, W., & Wang, J.Z. (2024). Attention-based CNN-LSTM for high-frequency multiple cryptocurrency trend prediction. *Expert Systems with Applications*, 237, 121520.
- Prabakaran, N., Palaniappan, R., Kannadasan, R., Dudi, S.V., & Sasidhar, V. (2021). Forecasting the momentum using customised loss function for financial series. *International Journal of Intelligent Computing and Cybernetics*, 14(4), 702-713.

- Ranjan, S., Kayal, P., & Saraf, M. (2023). Bitcoin price prediction: a machine learning sample dimension approach, *Computational Economics*, 61(4), 1617-1636.
- Saraswathi, S., Elavazhagan, A., Sabharwal, J.S., & Mohammad, S.I. (2024). An ensemble machine learning technique for bitcoin price prediction. *Journal of Trends in Computer Science and Smart Technology*, 6(2), 153-167.
- Singh, U., Tamrakar, S., Saurabh, K., Vyas, R., & Vyas, O.P. (2024). Optimizing hyperparameters of deep learning models for stock price prediction. In *2024 15th International Conference on Computing Communication and Networking Technologies* (pp. 1-7). IEEE. Kamand, India.
- Tanwar, S., Patel, N.P., Patel, S.N., Patel, J.R., Sharma, G., & Davidson, I.E. (2021). Deep learning-based cryptocurrency price prediction scheme with inter-dependent relations. *IEEE Access*, 9, 138633-138646.
- Wang, J., Feng, K., & Qiao, G. (2024a/b). A hybrid deep learning model for Bitcoin price prediction: data decomposition and feature selection. *Applied Economics*, 56(53), 6890-6905.
- Zhao, Y., Zhang, W., & Liu, X. (2024). Grid search with a weighted error function: hyper-parameter optimization for financial time series forecasting. *Applied Soft Computing*, 154, 111362.
- Zhong, C., Du, W., Xu, W., Huang, Q., Zhao, Y., & Wang, M. (2023). LSTM-ReGAT: a network-centric approach for cryptocurrency price trend prediction. *Decision Support Systems*, 169, 113955.

Original content of this work is copyright © Ram Arti Publishers. Uses under the Creative Commons Attribution 4.0 International (CC BY 4.0) license at <https://creativecommons.org/licenses/by/4.0/>

Publisher's Note- Ram Arti Publishers remains neutral regarding jurisdictional claims in published maps and institutional affiliations.